

# How Grammatical Gender Affects Perspective Taking

**Elena Andonova (eandonova@nbu.bg)**

Department of Cognitive Science and Psychology, 21 Montevideo Street  
Sofia, 1618, Bulgaria

**Zornitsa Savcheva (zsavcheva@gmail.com)**

Department of Cognitive Science and Psychology, 21 Montevideo Street  
Sofia, 1618, Bulgaria

**Gergana Todorova (gergana.tv@gmail.com)**

Department of Cognitive Science and Psychology, 21 Montevideo Street  
Sofia, 1618, Bulgaria

## Abstract

This is the first study to examine the influence of gender-sex congruence (match or mismatch between grammatical gender markers and participant sex) on the embodied processing of first-person sentences and images with either an internal or an external perspective in a picture-sentence verification task in Bulgarian modeled on Brunye et al.'s (2009) experimental paradigm. Participants were shown not to discriminate between perspectives when the grammatical gender was congruent with their own sex, thus allowing for an agentive interpretation by the reader. However, in the gender-sex incongruent condition, a significant 83 ms effect of image perspective was observed indicating large processing costs for attempting to adopt an internal perspective when the participant's sex was incompatible with the first-person gender marking, hence with action simulation from an egocentric perspective. These results are discussed in terms of embodiment specificity accounts and the experiential basis of grammar processing.

**Keywords:** perspective; grammatical gender; embodiment; Bulgarian.

## Introduction

There has been growing consensus in recent years that mental simulation of depicted events is an integral part of the process of reading. Although simulation may occur more naturally when the simulating agent and the narrative protagonist (actor) share certain identifying characteristics, it has been argued that readers may in fact embody various perspectives associated with multiple characters and bring this knowledge together in the build-up and upkeep of a relevant 'situation model' implied by narrative text (Zwaan & Radvansky, 1998). Even though it may be tempting to expect that the first-person (egocentric simulation) perspective on events would be the preferred default mode, this is by no means the general case. Even if told in the first-person singular, narratives from the viewpoint of characters with whom readers are reluctant to, or have difficulties to, identify may turn out to produce the opposite effect of distancing or misaligning oneself with a character.

Furthermore, although mental simulation may be a mechanism underlying the embodiment of abstract meaning, the latter is *per se* not a universal abstraction but has been shown to be both body-specific and culture-specific. A series of studies (e.g., Casasanto, 2009, Casasanto, 2011, Willems, Hagoort, & Casasanto, 2010) have provided support for the so-called *body-specificity hypothesis* of embodiment, including evidence that right-handers, on the one hand, and left-handers, on the other, represent abstract ideas with a positive and negative valence respectively, in spatially divergent directions of association, in thought, speech, and gesture. These two groups were also found to use correspondingly different areas of the brain for imagining actions and representing the meanings of action verbs. In fact, in line with the evidence accumulated so far, this *hypothesis* has been upgraded to a *theory of bodily relativity* (Casasanto, 2011, Casasanto, 2014).

It is not only bodies and their specific experiences that may differ but also cultural traditions of entire communities. In Western culture, for example, nodding is commonly associated with agreement and shaking one's head is interpreted as a sign of disagreement. There are, however, exceptions to this common pattern of association of head movement direction with acceptance vs. rejection, for example, the cultural convention in Bulgaria is the use of vertical head movement to mean 'No' and side-to-side head movement to mean 'Yes.' Both movement types are also generally slower, especially the side-to-side affirmative gesture, in comparison with Westerners' rather brisk lateral shake gesture. Thus, these two cultures spatially "embody" agreement via different movement patterns. Such cross-cultural habitual bodily movement differences may even affect certain aspects of cognitive processing that have no communicative intent (Andonova & Taylor, 2012). These findings speak in favor of the need to consider the ways in which embodiment may be grounded in culturally specific experiences of bodily associations with abstract thought.

If speakers and readers process verbal information in a way that is based on mental simulation of bodily experience, then it is reasonable to assume that differences in the perspective implied by the text from which the situation

model is constructed could also affect text comprehension. In particular, the use of first-person vs. second- or third-person perspective as implied by the personal pronoun may lead to differences in the relative ease of comprehension. Brunye, Ditman, Mahoney, Augustyn, and Taylor (2009) have demonstrated that pronoun variation and discourse context mediate the degree of embodiment experienced during narrative comprehension as measured by ease of simulation. Participants in their first experiment read simple sentences of the kind “*I am slicing the tomato*” where the personal pronoun varied between *I*, *you*, and *he* (first-, second-, and third-person) and saw images that either depicted a matching action-and-object combination or not. In this picture-sentence verification task, participants were faster to respond correctly to internal rather than external perspective images following the use of both first-person and second-person pronouns, although the former effect was only found in the participants analysis and not in the item analysis. The opposite response latency pattern emerged with third-person pronouns, i.e., faster verification for external than for internal-perspective pictures. However, previous studies using single-sentence stimuli (Borghi et al., 2004; Brunye et al., 2009) have shown that although second-person pronouns consistently cued an embodied agency perspective, first-person pronouns have only been found to do so inconsistently. In their second experiment, by introducing a two-sentence context to the protagonist, Brunye et al. (2009) succeeded in showing that first-person perspective in text is not associated with agency by participants when context does not encourage them to identify with a given character, e.g. *I am a deli employee*. In such a case, participants performed the picture-sentence verification task faster when they were processing the external than the internal image perspective.

Unlike English, many of the world’s languages, from Albanian in Albania to Zazaki in Turkey, employ grammatical gender categories to track down referents in discourse. Thus, in some of these languages, spoken or written discourse offers another cue to speaker/writer identity via the use of a grammatical gender marker on predicates in syntactic agreement with the first-person pronoun *I*. Consider the following examples in Bulgarian:

- (1) *Оцветил съм рибката.*
- (2) *Оцветила съм рибката.*

For both examples (1) and (2), the appropriate translation into English would be ‘*I have colored in the fish.*’ The first-person perspective is encoded in the auxiliary verb form *съм*. However, the past participle which is part of the predicate “colored-in” is either in the Masculine gender in example (1) or in the Feminine gender in example (2). Thus, a woman speaking a Slavic language (e.g., Bulgarian) would have to use feminine gender forms as in (2), and men would express reference to their first-person perspective by using masculine gender verb forms as in (1). This cue to referent identity is strong enough without additional context.

In order to investigate further the effects of first-person reference on adopting an embodied agency perspective in sentence interpretation, we designed an experimental study that manipulated this particular feature of grammatical gender in Bulgarian in combination with variation in image perspective (internal vs. external). As previous studies have shown more consistent results for the use of second-person pronouns (Brunye et al., 2009), we focused on first-person reference forms exclusively given that they may be variably associated with an internal (embodying) or external agency interpretation.

We expected to find little or no difference in the accuracy scores as picture-sentence verification is a relatively easy task. However, we predicted variation in response latencies as a function of experimental condition.

If reading first-person sentence descriptions activates an egocentric (reader’s) perspective rather than an allocentric (writer’s) perspective, then overall responses to internal perspective images should be faster than responses to external perspective images. This is what Brunye et al. (2009) established in their first experiment.

In addition, assuming an egocentric bias in first-person sentence interpretation, responses in the gender-sex incongruent condition should be slower than those in the gender-sex congruent condition. For example, adopting an egocentric perspective would be easier for women when the first-person reference was to an action performed by a woman as encoded in the gender marker of the predicate and much more problematic if the marker was of the masculine gender leading to an interpretation of a male actor. The opposite pattern would be the case with male readers of sentences marked for the feminine vs. the masculine gender. It was less clear in advance, however, if these two effects would be independent (additive) or interacting (super-additive), if each would emerge in the presence of the other. A super-additive effect would manifest itself statistically in the form of an interaction, since the responses to each of the two levels of one of the factors would depend on the level of the other factor. Generally speaking, we expected an interaction between the two variables on response times.

Alternatively, if reading first-person sentences does not automatically trigger an embodied agency interpretation but allows for multiple viewpoints to be entertained more readily and simultaneously, i.e., both reader/listener and writer/speaker as the actor, then these effects should not emerge as one would expect no cognitive effort to be spent on switching between perspectives or choosing between them.

## Method

The study implemented a 2 x 2 experimental design with Grammatical Gender - Sex Congruence (congruent vs. incongruent) and Perspective (internal vs. external) as within participant factors.

## Participants

48 participants (14 men and 34 women) took part in the experiment, mean age 23.27 years. They were university students within the 19-35 age range who volunteered and/or participated in exchange for course credit. All were native speakers of Bulgarian except two whose data were dropped. Given that the procedure required a right-hand response and speed of processing was measured, the data of three left-handed participants (two men, one woman) and one ambidextrous participant (male) were also excluded. The data of the remaining 42 participants (11 men and 31 women) were included in the analyses.

## Stimuli

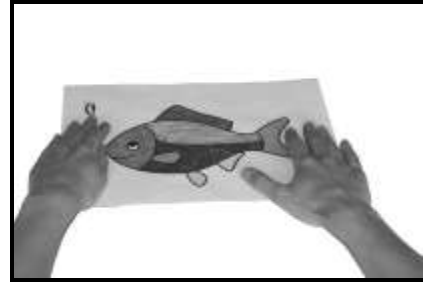
72 sentences (32 targets, 32 fillers and 8 practice sentences) describing simple completed actions were constructed in Bulgarian such as each sentence included the past participle of the corresponding verb (e.g., sliced, opened, etc.), the first-person form of the auxiliary verb 'be' (*сѣм*) and a direct object (e.g., tomato, can, box). Furthermore, each sentence had two forms – one in the masculine gender and one in the feminine, as indicated by the form of the past participle of the main verb. Actions described by the sentences were chosen in such a way so that they could not be readily identified as something typically performed by men or women in particular.

Two pictures for each of the 72 action descriptions were created – one presenting the action scene from an internal perspective and the other presenting it from an external perspective. Pictures corresponding to target sentences always depicted the completed action on the object as described in the sentence. Pictures used with filler sentences depicted either the same action (though performed on a different object), or the same object (though subjected to a different action) but never a truly matching action-object combination. In this respect, half of the practice pictures were similar to target pictures and the other half were similar to filler pictures.

All pictures were taken from the same viewing distance (ca. 92 cm) and the same downward angle (ca. 30°). They were all subsequently turned to grayscale and pasted onto a white background.



(a) external perspective



(b) internal perspective

Figure 1: Examples of image stimuli taken from the internal (b) and external (a) perspectives.

## Procedure

The experiment was conducted in sound-attenuating booths. On a computer screen participants read sentences describing a completed action after which they saw a picture depicting the same or different action and/or object. Their task was to verify, as quickly and accurately as possible, whether the sentence described the completed action depicted in the picture or not by pressing one of the two response keys on a button box. Participants responded with the index finger of their right hand. Response keys' associations with a positive or negative response were counterbalanced across participants.

The practice session consisted of 8 pseudorandomized trials. In the experimental session, the 64 trials were also pseudorandomized, such as each of the four conditions contained an equal number of stimuli. Across participants each stimulus appeared in all possible factor combinations. Target trials always required a YES response while filler trials always required a NO response.

Each trial in the practice and the experimental sessions contained the following sequence of displays: a fixation cross appeared for 400 ms followed by a sentence which remained on screen for 3000 ms, then another fixation cross appeared for 400 ms preceding a picture which was presented on the screen until a response was registered but for no more than 2000 ms. An inter trial interval of 1500 ms separated the distinctive trials.

## Results

Out of the set of thirty-two items, two items associated with technical errors in stimuli presentation were removed from the data under analysis.

Correct responses to target stimuli were associated with pressing the Yes button to verify a match between the meaning of the sentence and the action depicted in the photograph. Accuracy was calculated on the basis of these responses. Reaction times were registered for all responses but were only analyzed for correct responses to target stimuli. An equivalent number of filler trials were also included in order to balance for Yes and No responses.

Filler trials consisted of sentence-picture mismatches and the correct response in these cases was always negative.

### Response Accuracy

Overall, participants were highly accurate on the picture-sentence verification task ( $M=94.20\%$   $SD=9.00\%$ ).

Accuracy of response data was analyzed in a 2 (Gender-Sex Congruence: congruent vs. incongruent) x 2 (Perspective: internal vs. external) repeated measures ANOVA on participant means for the mean proportion of correct responses.

There were no main effects of gender-sex congruence or of perspective but we found a marginal two-way interaction between these variables,  $F(2, 41) = 3.38, p = .073, \eta_p^2 = .076$ . Whereas picture-sentence verification judgments were very similar in the internal picture perspective, participants' accuracy differed in the external perspective condition in which their judgment accuracy was higher in those cases when the grammatical gender of the predicate corresponded to the participant's sex, i.e., in cases of gender-sex congruence (see Table 1).

Table 1: Mean participant accuracy in percentages for target picture-sentence verification per condition.

Condition	Mean % (SD)
Congruent Internal	93.32 (10.68)
Congruent External	96.56 (7.72)
Incongruent Internal	94.34 (9.98)
Incongruent External	92.56 (9.08)

A 2 (Gender-Sex Congruence: congruent vs. incongruent) x 2 (Perspective: internal vs. external) repeated measures analysis of variance on item means for the proportion of correct responses yielded no statistically significant results.

### Response Latency

Participants responded with an average response latency of 1052 ms ( $SD=32$  ms) in this task. Only response latencies for correct Yes responses to target picture-sentence stimuli were analyzed.

A 2 (Gender-Sex Congruence: congruent vs. incongruent) x 2 (Perspective: internal vs. external) repeated measures analysis of variance on participant means for response times produced a main effect of perspective ( $F(1, 41) = 5.98, p = .019, \eta_p^2 = .127$ ) and no main effect of congruence. However, there was a significant two-way interaction between congruence and perspective, ( $F(1, 41) = 4.23, p = .046, \eta_p^2 = .094$ ).

The two-way interaction revealed that participants' correct responses did not differ in terms of speed when the predicate's grammatical gender was congruent with

participant sex, i.e., when women read sentences where the gender-marked verb form was consistent with a female agent and when men read sentences when the gender-marked form was masculine, thus consistent with a male agent of the action described. However, in the incongruent gender-sex condition, response times were significantly slower when the task required verification of visual stimuli in the internal than in the external perspective. In fact, there was an impressive 83 ms difference between participant means in these two situations.

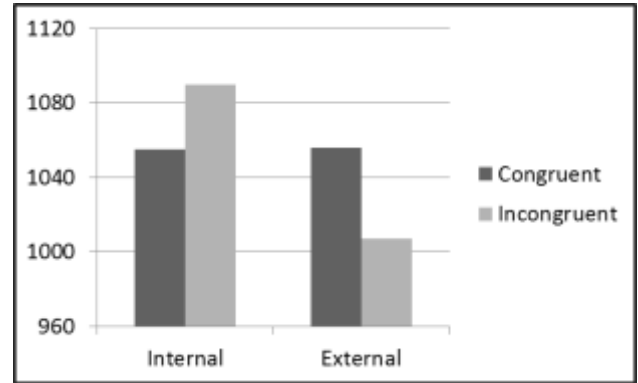


Figure 2: Mean participant response times (ms) to gender-sex congruent vs. incongruent picture-sentence stimuli shown in an internal vs. external perspective.

A 2 (Gender-Sex Congruence: congruent vs. incongruent) x 2 (Perspective: internal vs. external) repeated measures analysis of variance on item means for response times yielded similar statistical results. We found a marginal main effect of perspective ( $F(1, 29) = 4.19, p = .050, \eta_p^2 = .126$ ), no main effect of congruence, and a marginally significant two-way interaction between congruence and perspective, ( $F(1, 29) = 3.98, p = .056, \eta_p^2 = .121$ ). These results followed the same pattern as the findings from the analysis of participant means.

## Discussion

In this 2 x 2 experimental study on the influence of grammatical gender on perspective taking in reading simple first-person singular action sentences in Bulgarian, we found an interaction between gender-sex congruence and image perspective on processing times. Image perspective varied between internal, from the viewpoint of the reader looking at the screen, and external, from a 180° rotated viewpoint. Gender-sex congruence was a factor that encoded whether the grammatical gender of the predicate (Masculine vs. Feminine) corresponded to the participant's sex (male vs. female, respectively). In comprehension of a first-person (pronoun) sentence, the existence of such a correspondence (gender-sex congruence) allows for an agentive interpretation from the perspective of the reader,

that is, the reader can identify himself or herself with the doer of the action described in the sentence. Note, however, that such a correspondence works as a reliable cue to the intended referent only in situations where the reader's sex is uniquely congruent with the grammatical gender marker. If both speaker and hearer, or writer and reader, are of the same sex, then an ambiguity arises as to which of the two conversational partners is uniquely intended as the referent since the grammatical marker matches the sex of both partners. If the two interlocutors differ in terms of their sex, then only one of them could be seen as the doer of the action. In the process of reading, however, it is often the case that the reader is unaware of the writer's sex, thus producing greater uncertainty in interpretation.

We found that image perspective and gender-sex congruence had an interactive influence on verification processing times and even, though less so, on verification accuracy. In the gender-sex congruence condition, participants verified picture-sentence combinations equally fast in both internal and external perspectives. This result could be interpreted as evidence that multiple perspectives are simultaneously present in sentence comprehension in the absence of a uniquely identifiable referent. However, when grammatical gender markers and participant sex were not a fit, then the mismatching grammatical gender cue could clearly indicate that the first-person reference in such sentences excluded an embodied agency interpretation on behalf of readers and only allowed for an interpretation of the writer of these first-person sentences as the agent of the completed action described in them. As a result of this, performing picture-sentence verification on sentences indicative of a non-participant agent in combination with a participant aligned image internal perspective led to higher processing costs as seen in longer response times in comparison with an image external perspective.

The pattern of results from this study provides an interesting contrast to some previous findings described in the literature. For example, unlike Brunye et al.'s (2009) English-speaking participants, the Bulgarian speakers here showed no difference between processing internal and external perspective images when paired up with sentences where the first-person reference was gender-sex congruent. In Brunye et al.'s first experiment, participants responded faster to internal than to external perspective stimuli out of context allowing the authors to conclude that readers had adopted an egocentric perspective in sentence interpretation. The finding from our study is that there was no processing cost, hence no evidence of a preferred or default perspective taken in the interpretation of sentences, either egocentric or allocentric. It appears that in the gender-sex congruent condition readers were equally at ease with either interpretation. This result is in line with expectations that perspective may be ambiguous or uncertain in sentence comprehension unless there is a clear gender/sex difference in conversational roles between the reader and writer of the discourse element. It is also in line with Franklin, Tversky, and Coon's (1992) conclusion that speakers comprehending

narrative descriptions of various spatial scenes and multiple viewpoints on them seemed to take a neutral perspective when there was more than one probed point of view, rather than switch perspectives.

It could be argued that the gender-sex congruent condition in our study is the one that comes closer to the stimuli in Brunye et al.'s study. The gender-sex incongruent condition is clearly irrelevant in terms of a comparison with English language sentences where predicate gender is not a grammatical feature. It may, however, be difficult to draw such a comparison, even with the congruent condition. In one sense, it is neutral or balanced in that it does not preclude agency to be attributed to one of the conversational role partners or the other. In a different sense, however, the very existence of grammatical gender concord in a language may predispose its speakers to pay extra attention to this feature and to its extra-linguistic reference. Consider, for example, recent research on speakers of German, another language with a three-gender grammatical category on nouns. Imai, Schalk, Saalbach, and Okada (2010) found that German speakers projected sex onto grammatical gender and often made erroneous inferences in line with this projection. Similarly, German-speaking children appeared to rely on grammatical gender as a cue when determining whether a general biological property applies to a given object (Saalbach, Imai, & Schalk, 2012). In addition, studies on grammatical gender processing in Bulgarian and Italian speakers have revealed that men and women exhibited differential sensitivity to masculine and feminine gender word forms (Andonova, D'Amico, Devescovi, & Bates, 2004, Andonova, 2013). Whether or not the presence in a language of a grammatical category such as gender that serves a referent indexing function produces a different set of constraints on processing perspective in comparison with gender non-marking languages is a question that can better be answered in future cross-linguistic research. Alternatively, it can also be addressed in studies on processing perspective-related sentences within a gender marking language such as Bulgarian by introducing a comparison between gender-marked predicates (as in this study) and gender-neutral predicate forms, for example, describing actions in the Present tense which may serve as a baseline.

It is important to point out here that the gender-sex congruence by perspective interaction was obtained in our study for first-person sentences describing not currently depicted actions but completed actions, i.e., actions that had brought about a result seen in a visually presented scene. This distinction is important because our results highlight speakers' embodied attention to congruently described action outcomes and not only currently executed actions, that is, a certain level of abstraction away from the sensorimotor grounding of directly embodied action. However, if the study had produced a main effect of congruence and no interaction with perspective, then the interpretation of the results would have been constrained to an effect of speakers' attention to gender-sex congruence as a cue to

referent identity. The emergence of an interaction with perspective, however, speaks in favor of an embodied agency interpretation to these simple first-person sentences.

The interaction outcome is important for another reason as well. A simple main effect of perspective in this task may otherwise be due to higher visual processing costs of images presented in the external perspective. Without an interaction of perspective with a second variable, this possibility would be hard to discard. In Brunye et al.'s (2009) study, additional variables of interest were the manipulation of first- vs. second- and third-person sentential materials and context. In our study, the perspective manipulation alone would not have been sufficient to imply an embodied interpretation but the interaction between image perspective and gender-sex congruence enables such a reading of the results.

### Conclusion

In sum, this is the first study to show effects of grammar on perspective taking in comprehending simple action sentences. The grammatical feature that produces an effect here is grammatical gender, a widespread linguistic phenomenon (Corbett, 1991, 2005). The grammatical gender associated with the subject of the sentence in agreement with the predicate serves as a cue to attributing agency to a referent, in the case of first-person narratives, to the writer or potentially the reader of the sentence. While a correspondence between first-person gender and participant sex does not uniquely identify the reader as the intended referent, a mismatch between the two is sufficient to signal that the reader *cannot* be the agent of the completed action described in the sentence and thus to induce external perspective taking in the comprehension process.

It is common to assume that specificity in embodied processing, as in the differences between left-handers and right-handers, for example, comes about as a result of having performed actions in systematically different ways, although some manipulations such as wearing a cumbersome glove in a motor-fluency task can reverse the effects turning right-handers temporarily into “left-handers” (Willems, Hagoort, & Casasanto, 2010, Casasanto & Chrysikou, 2011). The alignment or misalignment of grammatical gender with self-reference is also the product of a lifetime of experience in a language community and cultural context. Unlike the reversibility of effects shown with handedness, however, these associations are difficult to override. They leave a trace on every aspect of language processing that embeds grammatical gender reference to extra-linguistic reality. The experiential basis of grammar processing and its role in perspective taking and embodiment are promising directions for future research.

### Acknowledgment

We would like to thank four anonymous reviewers for their helpful comments and suggestions on an earlier draft.

### References

- Andonova, E. (2013). Gender and Sex: The Experiential Basis of Grammar. In M. Knauff, M. Pauen, N. Sebanz, & I. Wachsmuth (Eds.), *Proceedings of the 35th Annual Conference of the Cognitive Science Society* (pp. 1738–1743). Austin, TX: Cognitive Science Society.
- Andonova, E., D’Amico, S., Devescovi, A., & Bates, E. (2004). Gender and lexical access in Bulgarian. *Perception & Psychophysics*, 66(3), 496-507.
- Borghi, A. M., Glenberg, A. M., & Kaschak, M. P. (2004). Putting words in perspective. *Memory & Cognition*, 32(6), 863-873.
- Brunyé, T. T., Ditman, T., Mahoney, C. R., Augustyn, J. S., & Taylor, H. A. (2009). When you and I share perspectives pronouns modulate perspective taking during narrative comprehension. *Psychological Science*, 20(1), 27-32.
- Casasanto, D. (2009). Embodiment of abstract concepts: good and bad in right- and left-handers. *Journal of Experimental Psychology: General*, 138(3):351–367.
- Casasanto, D. (2011). Different bodies, different minds: the body specificity of language and thought. *Current Directions in Psychological Science*, 20(6), 378-383.
- Casasanto, D. (2014). Bodily relativity. *Routledge handbook of embodied cognition*, Routledge, New York, 108-117.
- Casasanto, D., & Chrysikou, E. G. (2011). When left is “right” motor fluency shapes abstract concepts. *Psychological Science*, 22(4), 419-422
- Corbett, G. G. (1991). *Gender*. Cambridge: Cambridge University Press.
- Corbett, G. G. (2005). ‘The number of genders,’ ‘Sex-based and non-sex-based gender systems,’ and ‘Gender assignment systems’ [three chapters and maps]. In Haspelmath M, Dryer M, Gil D & Comrie B (eds.) *World atlas of language structures*. Oxford: Oxford University Press.
- Franklin, N., Tversky, B., & Coon, V. (1992). Switching points of view in spatial mental models. *Memory & Cognition*, 20(5), 507-518.
- Imai, M., Schalk, L., Saalbach, H., & Okada, H. (2010b). Influence of grammatical gender on deductive reasoning about sex-specific properties of animals. In R. Catrambone & S. Ohlsson (Eds.), *Proceedings of the 32<sup>nd</sup> annual conference of the cognitive science society* (pp. 1160–1165). Austin, TX: Cognitive Science Society.
- Saalbach, H., Imai, M., & Schalk, L. (2012). Grammatical Gender and Inferences About Biological Properties in German-Speaking Children. *Cognitive Science*, 36(7), 1251-1267.
- Willems, R. M., Hagoort, P., & Casasanto, D. (2010). Body-specific representations of action verbs neural evidence from right- and left-handers. *Psychological Science*, 21(1), 67-74.
- Zwaan, R. A., & Radvansky, G. A. (1998). Situation models in language comprehension and memory. *Psychological bulletin*, 123(2), 162.